Posture Estimation for Improved Photogrammetric Localization of Pedestrians in Monocular Infrared Imagery

Mikolaj E. Kundegorski, Toby P. Breckon
School of Engineering and Computing Sciences, Durham University, UK

ABSTRACT

Target tracking complexity within conventional video imagery can be fundamentally attributed to the ambiguity associated with actual 3D scene position of a given tracked object in relation to its observed position in 2D image space. Recent work, within thermal-band infrared imagery, has tackled this challenge head on by returning to classical photogrammetry as a means of recovering the true 3D position of pedestrian targets. A key limitation in such approaches is the assumption of posture – that the observed pedestrian is at full height stance within the scene. Whilst prior work has shown the effects of statistical height variation to be negligible, variations in the posture of the target may still pose a significant source of potential error. Here we present a method that addresses this issue via the use of Support Vector Machine (SVM) regression based pedestrian posture estimation operating on Histogram of Orientated Gradient (HOG) feature descriptors. Within an existing tracking framework, we demonstrate improved target localization that is independent of variations in target posture (i.e. behaviour) and within the statistical error bounds of prior work for pedestrian height posture varying from 0.4-2.4m over a distance to target range of 7-30m.

Keywords: thermal target tracking, temporal filtering, intelligent target reporting, thermal imaging, pedestrian detection, people detection, sensor networks, temporal fusion, passive target positioning, 3D pedestrian localization

1. INTRODUCTION

Target tracking within conventional video imagery poses a significant challenge that is increasingly being addressed via complex algorithmic solutions. The complexity of this problem can be fundamentally attributed to the ambiguity associated with actual 3D scene position of a given tracked object in relation to its observed position in 2D image space. Recent work has tackled this challenge directly by returning to classical photogrammetry, within the context of current target detection and classification techniques, as a means of recovering the true 3D position of pedestrian targets within the bounds of current accuracy norms [1]. However, a key limitation in such approaches is the assumption of posture - that the observed pedestrian is at full height stance within the scene. Whilst prior work has shown the effects of statistical height variation to be negligible [1], variations in the posture of the target may still pose a significant source of potential error (see Figure 1). A non-cooperative pedestrian target may use variations within their posture to subvert accurate localization of their position within the scene (e.g. crawling or crouching). Despite this issue, in many applications the upright stance of a pedestrian target is indeed assumed [1–4].

Within the context of pedestrian tracking, our prior work in [1] demonstrated that reasonable performance can practically be achieved through the combined use of infrared imagery (thermal-band, spectral range: 8-12μm) and the application of real-time photogrammetry. A key advantage of such thermal-band infrared (IR) imagery for pedestrian localization is both andro-bust detection of human shape signatures within the scene [5–7] and robust localization of their scene bounds in pixel-space (e.g. Figure 1). As such, the principles of photogrammetry can be used to recover 3D pedestrian position within the scene based on a known camera projection model and an assumption that variance in human height is in fact quite small (statistically supported by [8, 9]). In [1] we experimentally investigated the accuracy of classical photogrammetry, within the context of current target detection and classification techniques [5–7], as a means of recovering the true 3D position of pedestrian targets within the scene. A real-time approach for the detection, classification and localization of pedestrian targets via thermal-band (infrared) sensing was presented with supporting statistical evidence underpinning the key photogrammetric assumptions.

Overall, despite extensive work in ground-based sensor networks [10–13], the use of photogrammetry within this context has received only limited attention [1, 14, 15]. The visible-band work of [15] uses a similar approach within a Bayesian 3D tracking framework but does not explicitly address issues of accuracy or its use within a detection filtering framework [1]. In addition, some general scene understanding approaches have also used this principle to determine relative object dimensions and positions within the scene [16, 17] although al-

ternative approaches such as active sensing [18], structure from motion [19] and monocular depth recovery [20, 21] have become increasingly popular within this task of late.

Prior work explicitly dealing with thermal-band (IR) imagery within an automated surveillance context is presently largely focused upon pedestrian detection [3, 5, 7, 22, 23] and tracking [24, 25]. More recently extended studies have investigated the fundamentals of both background scene modeling [26, 27] and feature point descriptors [28] that commonly form the basis of many such techniques [3, 5]. Early thermal-band work by [14] proposes a shape driven methodology for posture estimation based on torso orientation and limb localization but does not relate directly to numerical recovery of relative height as we specifically address here.

The work presented in this paper is a direct extension of [1] that demonstrates photogrammetric pedestrian localization within thermal-band imagery incorporating a lightweight tracking solution akin to that of [6]. In [1] photogrammetric pedestrian target localization is presented to an accuracy significantly within the commonly regarded *"gold-standard"* of consumer-level Global Position System (GPS) positioning (typically $\pm 5m$ under ideal conditions [29]). This success is based on a) the key advantage of reduced pixel-space localization ambiguity within thermal-band infrared imagery [5] and b) recent statistical results that report narrow standard deviations within large-scale surveys of human height variation [8, 9]. Furthermore, it is achieved using solely passive sensing from a monocular infrared imaging camera, with no *a priori* environment calibration.

Building directly on this framework presented in [1], here we present a method that addresses the remaining issue of posture variation via the use of regression based pedestrian posture estimation. The posture of a pedestrian target detected within the scene is estimated as a percentage of full height (maximal posture) based on the use of a Histogram of Orientated Gradient (HOG) feature descriptor extracted from each detected scene target and the use of Support Vector Machine (SVM) based machine learning regression**.** In contrast to prior work in the field, we leverage the key advantages of thermal-band infra-red (IR) imagery for pedestrian localization with a tracking framework [6] and demonstrate robust target localization, independent of variations in target posture (i.e. behaviour), within the statistical error bounds outlined in [1]. This is demonstrated for variations in pedestrian target height, due to posture, ranging from 0.4-2.4m over a distance to target range of 7-30m (Figure 3). We show that the robust localization and foreground target separation, afforded via infrared imagery, that facilitates accurate 3D position estimation of targets to within the error bounds of conventional Global Position System (GPS) positioning [1] can be extended to maintain such accuracy bounds despite variations in target posture. Based on our improved photogrammetric estimation of target position, we then illustrate the efficiency of regular Kalman filter based tracking operating on actual 3D pedestrian scene trajectories.

2. PEDESTRIAN TARGET LOCALIZATION

We perform localization, and subsequent tracking in real-world 3D space (*"scene space"*), based on the initial detection (Section 2.1) and photogrammetric based localization (Section 2.2). This follows the approach outlined in the prior work of [1].

2.1 Pedestrian Detection

Our approach is illustrated against the backdrop of classical two stage automated visual surveillance [1]. First we detect initial candidate regions within the scene (Section 2.1.1), thus facilitating efficient feature extraction over isolated scene regions, to which an identified target type is assigned via secondary object classification (Section 2.1.2) [5].

2.1.1 Candidate Region Detection

In order to facilitate overall real-time performance, initial candidate region detection identifies isolated regions of interest within the scene facilitating localized feature extraction and classification. By leveraging the stationary position of our sensor, this is achieved using a combination of two adaptive background modeling approaches [30, 31] working in parallel to produce a single robust foreground model over varying environmental conditions and notably within varying ambient thermal/infrared illumination conditions within complex, cluttered environments.

Within the first model, a Mixture of Gaussian (MoG) based adaptive background model, each image pixel is modeled as a set of Gaussian distributions, commonly termed as a Gaussian mixture model, that capture both noise related and periodic (i.e. vibration, movement) changes in pixel intensity at each and every location within the image over time [30, 32]. This background model is adaptively updated with each frame received and each pixel is probabilistically evaluated as being either part of the scene foreground or background following this methodology. The second model comprises the

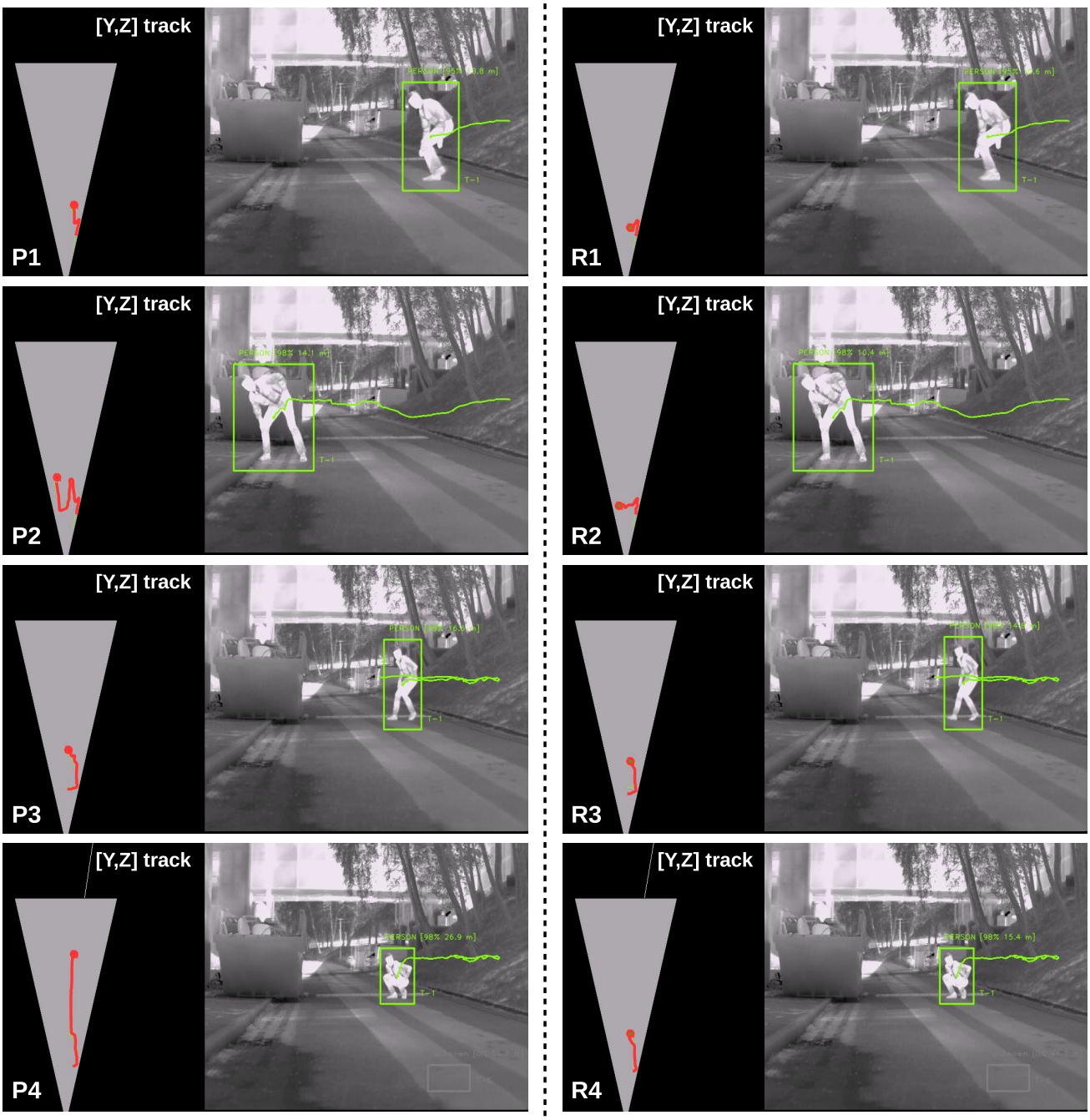

Figure 1. An example of real-time multiple pedestrian detection and tracking in infrared imagery with associated geo-referenced 3D tracks in the presence of posture variation - presented using basic photogrammetric position estimation (P1-P4 left, [1]) and with additional regression based posture estimation for localization correction (R1-R4 right, Section 2.2.2).

use of Bayesian classification in a closed feedback loop with Kalman filtered predictions of foreground component position [31]. Within this model, each pixel is similarly probabilistically classified as either foreground or background but this is further reinforced via Kalman predictions for the positions of foreground objects (i.e. connected component foreground regions [33]) present in the previous time-step. This object-aware model significantly aids in the recovery of fast moving foreground objects under varying illumination conditions such as the thermal gradients inherent within infrared imagery. Overall this combined approach provides a slowly-adapting background model in the traditional sense [30], that can be robust to rapid illumination gradients, whilst similarly providing foreground consistency to fast moving scene objects [31]. The binary output of each foreground, based on a probabilistic classification threshold, is combined conjunctively to provide robust detection of both static and active scene objects. For illustrative examples and further discussion the reader is directed to [1].

2.1.2 Pedestrian Classification

Subsequent classification follows the bag of visual words (or codebook) methodology [34–36] using Speeded Up Robust Features (SURF) approach [37] as our multi-dimensional features empirically suited towards thermal infrared imagery [1, 3, 5, 28]. Following this methodology, we perform feature extraction and clustering over all of the example training imagery (for all object classes) to produce a set of general feature clusters that characterise the overall feature space. Commonly this set of feature clusters is referred to as a codebook or vocabulary as it is subsequently used to encode the features detected on specific object instances (pedestrian or non-pedestrian) as fixed length vectors for input to both the initial off-line classifier training and on-line classification phase of such machine learning driven classification approaches. Here we perform clustering using the common-place k-means clustering algorithm in 128-dimensional space (i.e. SURF feature descriptor length of 128 [37]) into 1000 clusters. A given object instance is encoded as a fixed length vector based on the membership of the features detected within the object to a given feature cluster based on nearest neighbour (hard) cluster assignment. Essentially the original variable number of SURF features detected over each training image or candidate region is encoded as a fixed length histogram representing the membership of these features to each of these clusters. This fixed length distribution of features forms a feature vector that is then used to differentiate between positive and negative instances of a given class based on a trained classifier. The feature vector forms the input to a two-class SVM classifier, $pedestrian = \{yes, no\}$, that is trained using a RBF kernel, via grid-based kernel parameter optimization, within a cross-validation based training regime [38].

2.2 Photogrammetric Position Estimation

Firstly, we present a brief recap of our baseline localization approach as presented in [1] (Section 2.2.1) and subsequently show how this can be extended to address posture variation within detected pedestrian targets (Section 2.2.2).

2.2.1 Baseline Photogrammetry

Based on automated detection (Section 2.1.2), target position is initially known within *"sensor space"* (i.e. pixel position within the image). Consequently, target position is estimated based on the principles of photogrammetry together with knowledge of the perspective transform under which targets are imaged and an assumption on the physical (real-world) dimension of a target in one plane [1]. All targets are imaged under a standard perspective projection [33] as follows:

$$x = f\frac{X}{Z},\ y = f\frac{Y}{Z} \tag{1}$$

where real-world object position, (X, Y, Z), in 3D scene co-ordinate space is imaged at image pixel position, (x, y), in pixel co-ordinate space for a given camera focal length, f. We assume both positions are the centroid of the object with (x, y) being the centre of the bounding box, of the image sub-region, for a target (object) detected in the scene (Section 2.1.1, e.g. Figure 2).

With knowledge of the camera focal length, f, the original object (target) position, (X, Y, Z), can be recovered based on (assumed) knowledge of either object width, $\triangle X$, or object height, $\triangle Y$ (i.e. the difference in minimum and maximum positions in each of these dimensions for the object). From the bounds of the detected targets (Section 2.1.2) we can readily recover the corresponding object width, $\triangle x$, and object height, $\triangle y$, in the image. Based on this knowledge, rearranging and substituting into Eqn. 1 we can recover the depth (distance to target, Z) of the object position as follows:

$$Z = f'\frac{\triangle Y}{\triangle y} \tag{2}$$

Knowing Z via Eqn. 2, we can now substitute back into Eqn. 1 and with knowledge of the object centroid in the image, (x, y), we can recover both X and

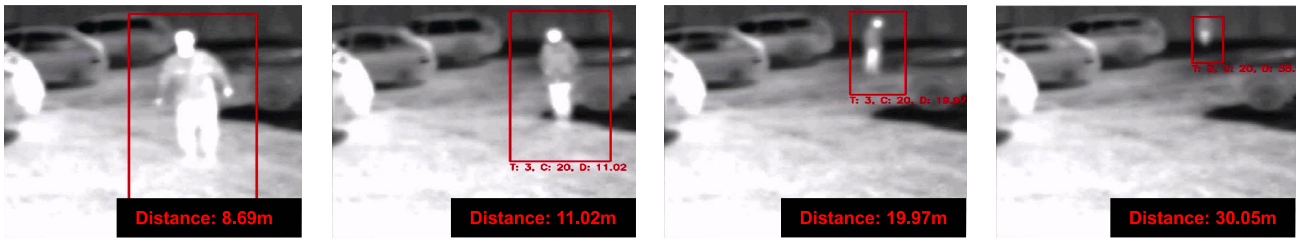

Figure 2. Photogrammetry facilitates the approximate recovery of a camera to target distance for an example target (person) without any need for additional (active) range sensing [5]

Y resulting in full recovery of real-world target position, (X, Y, Z), relative to the camera. In Eqn. 2, f' represents focal length, f, translated from standard units, mm, to focal length measured in pixels:-

$$f' = \frac{width_{image} \cdot f}{width_{sensor}} \tag{3}$$

where $width_{image}$ represents the width of the image (pixels), $width_{sensor}$ represents the camera CCD sensor width (mm).

Crucially, if we now assume a fixed width, $\triangle X$, or height, $\triangle Y$, for our object we can recover complete 3D scene position relative to the camera. For pedestrian detection we can assume average adult human height based on available medical statistics [8, 9]. Despite commonly held beliefs, notable large-scale studies have shown variance on human height within the adult population to be low ("*in populations of European descent, the average height is ∼178 cm for males and ∼165 cm for females, with a standard deviation of ∼7 cm*" [9]). As concluded in [1] this translates into a Z position error, attributable to height variation, at 60m distance is within GPS error tolerances ($\pm 5m$) for approximately the ∼2-98% percentile of the population (based on height distribution). The reader is directed to both the extensive statistical presentation and empirical verification of [1] for further discussion.

2.2.2 Addressing Posture Variation

Of course, a key limitation within this approach (Section 2.2.1) is variation in human height that is artificially introduced outside of this statistical variation - what if the pedestrian varies their posture such that their height does not conform to the statistical assumptions in use? In order to address this issue we use a regression based approach to map a dense gradient-based feature descriptor extracted from the image, itself capturing inherent body posture, to a numerical approximation of height relative to full upright posture.

Based on our detected pedestrian region (from Section 2.1), we thus extract Histogram of Oriented Gradient (HOG) features [4] as an input to SVM regressor. This predicts the numerical percentage of full human height, {0..100+}, represented by the current posture. The HOG descriptor is based on histograms of oriented gradient responses in a local region around a given pixel of interest. A rectangular block, pixel dimension $b \times b$, is first divided into $n \times n$ (sub-)cells and for each cell a histogram of gradient orientation is computed (quantised into H histogram bins for each cell, weighted by gradient magnitude). The histograms for all cells are then concatenated and normalised to represent the HOG descriptor as a vector, $\overrightarrow{v}_{HOG}$, for a given block (i.e. associated pixel location). For image gradient computation centred gradient filters $[-1, 0, 1]$ and $[-1, 0, 1]^T$ are used as per [4].

To construct our HOG descriptor, the localized pedestrian region is first zero-padded to form a square image region and subsequently re-sampled to a uniform 128×64 pixel image size, $(h \times w)$. We then compute the global HOG descriptor of this localized region using a block stride, $s = 8$ ($H = 9$, $n = 2$, $b = 16$ from [4]), to form the input to the SVM regressor (υ-support vector regression, [39]). Based on this 3780 dimensional vector, $\overrightarrow{v}_{HOG}$, (i.e. $H \times n^2 \times (\frac{h}{s} - 1) \times (\frac{w}{s} - 1)$) we train using both Radial Basis Function (RBF) and linear kernels, with grid-based kernel parameter optimization, within a cross-validation based training regime. Training is performed over ∼11,000 example images captured of 10 individuals at varying heights (crawling to stretching, ∼40-140% of full height based on [8, 9]) under varying environmental conditions over distances in the range 6-60m (Figure 3). This results in a SVM regression function capable of mapping the HOG feature descriptor representation to a numerical approximation of current posture as a percentage of full height, $f_{SVM}() : \overrightarrow{v}_{HOG} \rightarrow\sim \{40...140\}$. Examples of the training images used for this task are shown in Figure 3. Empirically we use an output range of

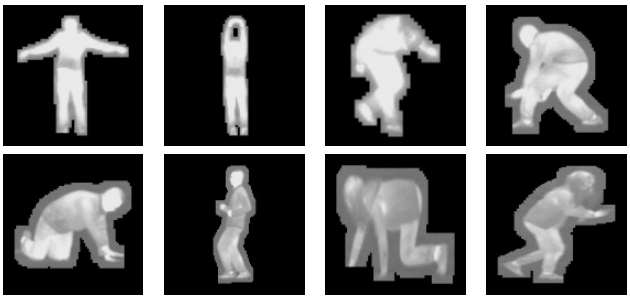

Figure 3. An illustrative subset of the training examples use for training the SVM regressor.

$f_{SVM}() \cong \{40...140\}$ to allow for pedestrian posture stretching beyond their head (e.g. in activities such as digging, climbing and alike, Figure 3).

This estimation of posture as a percentage of full height is then used as a scaling factor to adjust the pedestrian pixel height detected within the image, $\triangle y$, in order to compensate for posture within Eqn. 2 to recover target position such that $\triangle y' = \frac{100 \triangle y}{f_{SVM}()}$.

3. EVALUATION

Our results are presented using both quantitative measures of pedestrian localization accuracy (Table 1 / Figure 4) and qualitative assessment of 3D localization and tracking performance over a range of exemplar scenarios (Figures 1, 5, 6). All evaluation imagery is captured using an un-cooled infra-red camera (*Thermoteknix Miricle 307k*, spectral range: 8-12μm) with statistical performance measured using validation test set of 2,500 images drawn from the same variation and environmental conditions as used for training. Evaluation was performed around a variety of urban/industrial (cluttered) and suburban environments immediately around the School of Engineering and Computing Sciences, Durham University. Under evaluation conditions GPS accuracy locally was found to be $\pm 5m$, based on a consumer GPS unit [29].

Statistical variation is based on the coefficient of determination, r^2, that measures how well our given regression model, f_{SVM}, approximates a set of true data samples ($\{y_i\}$, i.e. the test set) with a result of $r^2 = 1$ indicating a perfect regressive fit to the data sample. The coefficient of determination is defined as the fractional remainder of the sum of squares of residuals (also called the residual sum of squares) over the total sum of squares (proportional to the variance of the data set). This is defined, as shown in Eqn. 4, for a given regressor function with predictive output, $f_i()$, for sample i against associated ground truth value, y_i, where $\bar{m}$ is the mean of the entire sample data set, $\{y_i\}$.

$$r^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{m})^2} \quad (4)$$

Here we have a SVM regressor function, $f_i() = f_{SVM}()$ and regressive target values, $y_i \cong \{40...140\}$, from which we calculate our r^2 statistics over the training set as presented in Table 1. In addition we present the mean estimation error (in % of full height, against ground truth data set $\{y_i\}$) and the associated standard deviation (Table 1). From the statistical evaluation of Table 1 we can see a strong regressive fit to the test sample data set based on the coefficient of determination outcome with the use of a RBF kernel with the SVM marginally outperforming the use of a linear kernel under the same conditions. In practice, the mean estimation error is also low for both kernel options ($\sim 3\%$) representing a real-world height estimation error of approximately 0.06 ± 0.04m on the average male height of 1.77m [9] (Table 1) which is within the error margins considered by [1] for reliable localization. The high standard deviation, compared to the mean, indicates high variability in the regressive estimation that is further illustrated by Figure 6.

The strong performance of the linear kernel, determining a decision boundary in the original feature space without the use of a projective kernel such as RBF, provides strong evidence that the 3780 dimension HOG descriptor indeed provides a good discriminator of human posture as a numerical percentage of full height. This general statistical evaluation (Table 1) is further supported by the specific statistical evaluation presented in the graph of Figure 4 where we again measure the coefficient of determination for a single individual performing various activities of varying posture over a range of (ground truth) distances in the range 7-28m. Figure 4 shows the stability of the numerical posture estimation approach over varying distance ranges.

This quantitative statistical evaluation (Table 1 / Figure 4) is further supported by the qualitative results

	coefficient of determination (r^2)	**mean estimation error (% of full height, $\pm\sigma$)**	**mean height error (on 1.77m male, $\pm\sigma$) (m)**
SVM (RBF kernel)	0.95	3.2±2.4	0.057±0.043
SVM (linear kernel)	0.93	3.8±3.0	0.067±0.053

Table 1. Statistical evaluation of pedestrian height regression accuracy (general)

presented in Figures1, 5 & 6. Figures 1 and 5 both illustrate a pedestrian tracking sequence using using standard photogrammetric position localization as per prior work [1] (P1-P4 left, Figures 1/5) and posture estimation via regression for localization correction as per Section 2.2.2 (R1-R4 right, Figures 1/5). These images are sequentially sub-sampled from the test scenarios with tracking and spatio-temporal detection performed as outlined in [1]. Within each sub-figure (Figures 1 & 5 R1-R4/P1-P4) we present the detected pedestrian(s) using a bounding box, associated 2D image projection of the track (P1-P4/R1-R4 insets, right) and the planar view of the $\{Y/Z\}$ tracked position relative to the camera (P1-P4/R1-R4 insets, left). Tracking and detection performance is as per [1].

From Figures 1 and 5 we can see that the accuracy and continuity of the $\{Y/Z\}$ position localization of the pedestrian from standard photogrammetric techniques [1] (shown in P1-P4 left, Figures 1/5) is significantly effected by changes in posture. Changes in posture in Figure 1 (e.g. transitions P1 $\rightarrow$P2 and P3 $\rightarrow$P4) show significant erroneous jumps in the spatial locality of pedestrian target when the planar view of the $\{Y/Z\}$ tracked position history is considered. This is similarly present in Figure 5 (e.g. transitions P1 $\rightarrow$P2 and into both of P3/P4) where again significant erroneous jumps in spatial locality are present despite the continuity of the target position relative to the camera within the scene. By contrast, with the use of posture estimation via regression to perform localization correction in both the Figure 1 and Figure 5 sequences we see a planar view of the $\{Y/Z\}$ tracked position history than remains consistent with the target position from the camera despite changes in posture. Our use of a regression based approach is shown to facilitate effective compensation for variations in target posture within photogrammetric pedestrian localization with (Figures 1/5).

Furthermore, Figure 6 shows an extended comparison of the resulting reported target position tracks as a planar view of the $\{Y/Z\}$ tracked position (for the scenario of Figure 5). These are reported relative to the camera position as per [1, 7]. This is illustrated using both standard photogrammetric position localization as per prior work [1] (Figure 6 A) and posture estimation via regression for localization correction as per Section 2.2.2 (Figure 6 B). As can be seen from Figure 6, without any compensation for variations in target posture we suffer significant erroneous jumps in the spatial locality of pedestrian target due to posture variation effecting the photogrammetric estimation of target position (Figure 6 A). By contrast, the use of posture estimation via regression facilitates recovery of a smooth track of the target motion (Figure 6 B) that is consistent with ground truth target motion within the scene (illustrated in sub-samples of Figure 5). The pedestrian target traverses away from the camera performing various (posture varying) activities and 15m, 25m and 35m marker points (Figure6 B).

4. CONCLUSIONS

Overall we have shown that the use of SVM driven regression facilitates effective posture estimation to enable improved 3D localization and tracking of pedestrians within infrared imagery based on the principles of photogrammetry. This directly advances the robustness of prior work in field for pedestrian localization in the presence of posture variation [1, 5]. Within the context of improved pedestrian localization in infrared thermal imagery and the use of the pedestrian tracking approach outlined in [1] significant improvements in spatial localization and hence track consistency are experimentally illustrated. These are further supported by a strong statistical evaluation of the chosen regression methodology.

This work further strengthens the application of pedestrian tracking within 3D scene-space that facilitates the ready disambiguation of multiple target tracking scenarios using low-complexity approaches with reduced computational overheads [1]. Our approach is demonstrated over multiple scenarios in cluttered environments where a clear improvement in tracking consistency is illustrated.

Future work will look to investigate the extension of this approach to the recovery of multiple pose attributes [14, 40], as an enabler to human activity classification [6, 41, 42] and also into visible-band imagery using recent advances in real-time salient

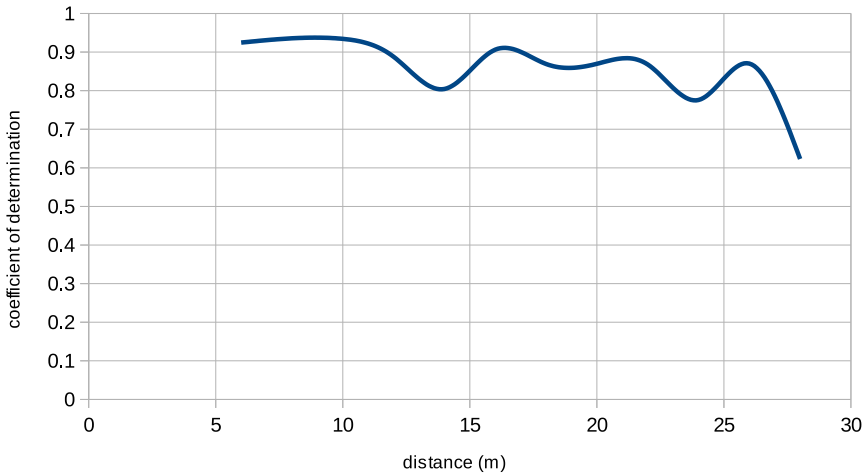

Figure 4. Statistical evaluation of pedestrian height regression accuracy against distance to target (specific, SVM with RBF kernel)

object detection [43]. Applicability within the context of mobile platform navigation [44–47], driver assistance systems [48–50] and for multi-platform, multi-modal wide-area search and surveillance tasks [7, 51, 52] will be further explored.

Acknowledgments: *This work was supported by the Defence Science and Technology Laboratory (UK MOD) and the Innovate UK. Supporting parts of this work, forming background material to the contribution made here, were originally published as part of [1, 7] by the same author/publisher.*

REFERENCES

1. M. Kundegorski and T. Breckon, "A photogrammetric approach for real-time 3d localization and tracking of pedestrians in monocular infrared imagery," in *Proc. SPIE Optics and Photonics for Counterterrorism, Crime Fighting and Defence*, vol. 9253, pp. 1–16, SPIE, September 2014.
2. P. Peng, Y. Tian, Y. Wang, J. Li, and T. Huang, "Robust multiple cameras pedestrian detection with multi-view Bayesian network," *Pattern Recognition*, vol. 48, pp. 1760–1772, May 2015.
3. B. Besbes, A. Rogozan, and A. Bensrhair, "Pedestrian recognition based on hierarchical codebook of SURF features in visible and infrared images," in *Proc. Intelligent Vehicles Symp.*, pp. 156–161, IEEE, June 2010.
4. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp. 886–893, 2005.
5. T. Breckon, J. Han, and J. Richardson, "Consistency in muti-modal automated target detection using temporally filtered reporting," in *Proc. SPIE Electro-Optical Remote Sensing, Photonic Technologies, and Applications VI*, vol. 8542, pp. 23:1–23:12, November 2012.
6. J. Han, A. Gaszczak, R. Maciol, S. Barnes, and T. Breckon, "Human pose classification within the context of near-ir imagery tracking," in *Proc. SPIE Optics and Photonics for Counterterrorism, Crime Fighting and Defence*, vol. 8901, pp. 1–10, SPIE, September 2013.
7. T. Breckon, A. Gaszczak, J. Han, M. Eichner, and S. Barnes, "Multi-modal target detection for autonomous wide area search and surveillance," in *Proc. SPIE Emerging Technologies in Security and Defence: Unmanned Sensor Systems*, vol. 8899, pp. 1–19, SPIE, September 2013.
8. R. Craig, J. Mindell, and V. Hirani, "Health survey for England," *Obesity and Other Risk Factors in Children. The Information Centre*, vol. 2, 2006.
9. P. M. Visscher, "Sizing up human height variation.," *Nature genetics*, vol. 40, pp. 489–90, May 2008.
10. A. Yilmaz, O. Javed, and M. Shah, "Object tracking - a survey," *ACM Computing Surveys*, vol. 38, pp. 13–es, Dec. 2006.
11. H. K. Aghajan and A. Cavallaro, *Multi-camera networks: principles and applications*. Academic press, 2009.
12. G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *J. of Ambient Intelligence and Humanized Comp.*, vol. 2, no. 2, pp. 127–151, 2011.
13. X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognition Letters*, 2012.
14. S. Iwasawa, K. Ebihara, J. Ohya, and S. Morishima, "Real-time estimation of human body posture from monocular thermal images," in *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp. 15–20, 1997.

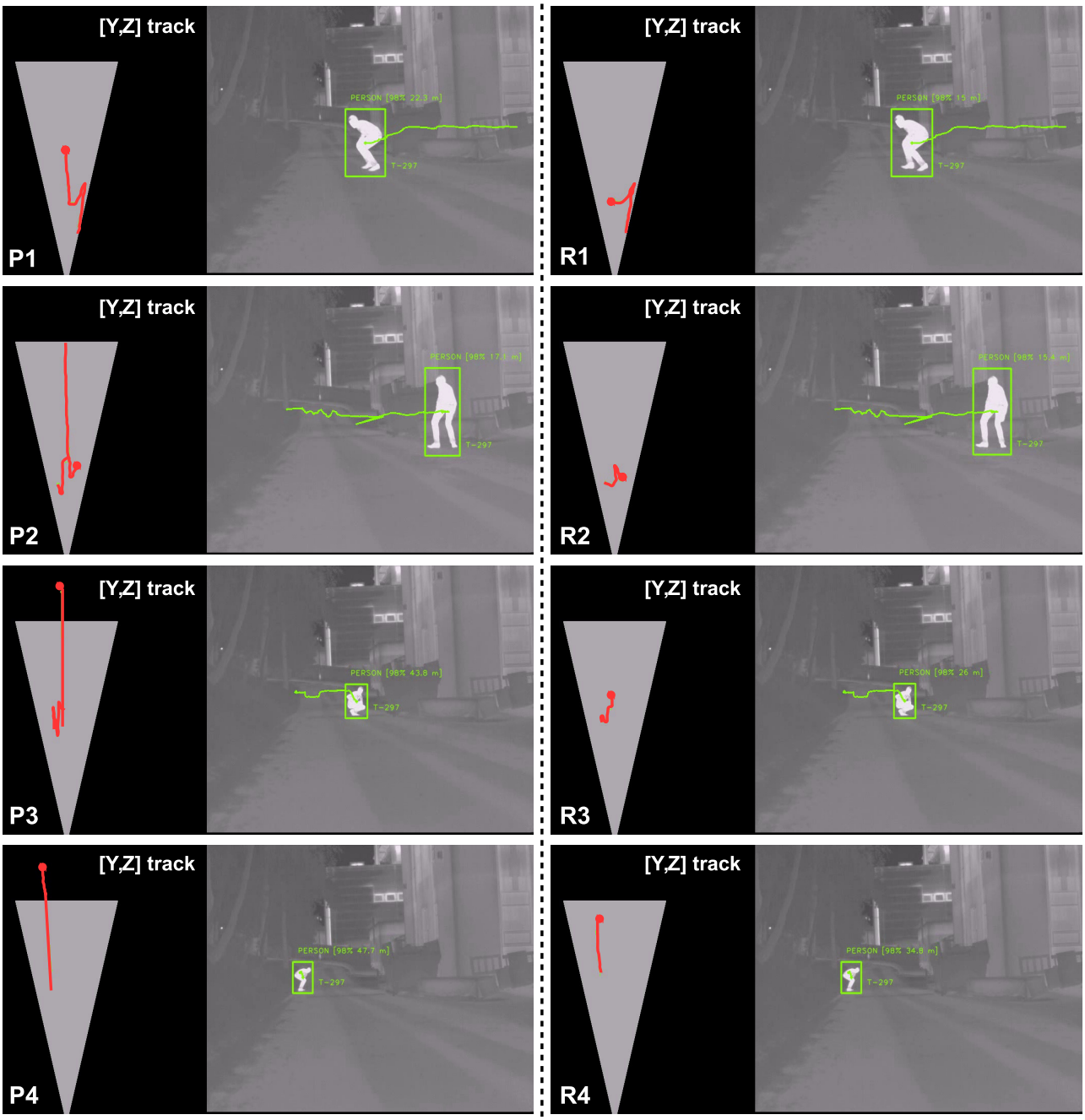

Figure 5. An example of real-time multiple pedestrian detection and tracking in infrared imagery with associated geo-referenced 3D tracks in the presence of posture variation - presented using basic photogrammetric position estimation (P1-P4 left, [1]) and with additional regression based posture estimation for localization correction (R1-R4 right, Section 2.2.2).

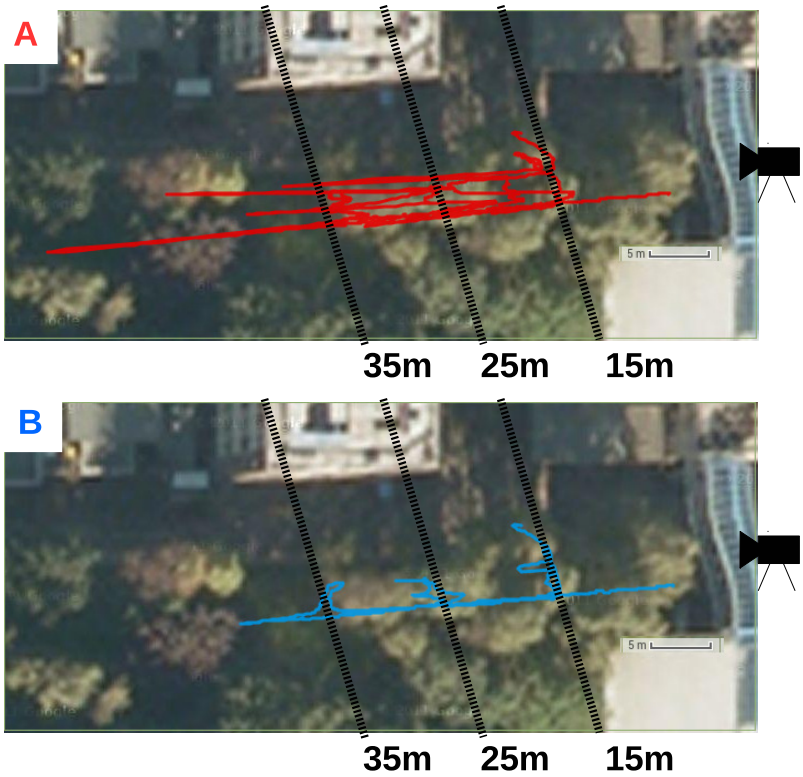

Figure 6. Exemplar track plotted on geo-referenced satellite imagery with corresponding ground truth distance ranges obtained from hand-held GPS tracking units - presented using basic photogrammetric position estimation (A upper, [1]) and with additional regression based posture estimation for localization correction (B lower, Section 2.2.2).

15. E. Brau, J. Guan, K. Simek, L. D. Pero, C. R. Dawson, and K. Barnard, "Bayesian 3D Tracking from Monocular Video," in *Int. Conf. Computer Vision*, pp. 3368–3375, 2013.
16. J.-F. Lalonde, D. Hoiem, A. A. Efros, C. Rother, J. Winn, and A. Criminisi, "Photo clip art," in *ACM SIGGRAPH*, vol. 26, p. 3, ACM Press, Aug. 2007.
17. J. Yuen, B. Russell, and A. Torralba, "LabelMe video: Building a video database with human annotations," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 1451–1458, IEEE, Sept. 2009.
18. G. Payen de La Garanderie and T. Breckon, "Improved depth recovery in consumer depth cameras via disparity space fusion within cross-spectral stereo," in *Proc. British Machine Vision Conference*, pp. 417.1–417.12, September 2014.
19. M. Lhuillier, "Incremental Fusion of Structure-from-Motion and GPS using Constrained Bundle Adjustments.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, July 2012.
20. D. Hoiem, A. Efros, and M. Hebert, "Geometric context from a single image," in *Int. Conf. Computer Vision*, vol. 1, pp. 654–661, IEEE, 2005.
21. Y. Diskin and V. K. Asari, "Dense 3D point-cloud model using optical flow for a monocular reconstruction system," in *Applied Imagery Pattern Recognition Workshop*, pp. 1–6, IEEE, Oct. 2013.
22. J. W. Davis and V. Sharma, "Robust detection of people in thermal imagery," in *Proc. Int. Conf. Pattern Recognition*, vol. 4, pp. 713–716, 2004.
23. J. W. Davis and V. Sharma, "Background-subtraction in thermal imagery using contour saliency," *Int. Journal of Computer Vision*, vol. 71, no. 2, pp. 161–181, 2007.
24. M. Yasuno, S. Ryousuke, N. Yasuda, and M. Aoki, "Pedestrian detection and tracking in far infrared images," in *Proc. Int. Conf. Intelligent Transportation Systems*, pp. 182–187, 2005.
25. J. Wang, D. Chen, H. Chen, and J. Yang, "On pedestrian detection and tracking in infrared videos," *Pattern Recognition Letters*, vol. 33, pp. 775–785, Apr. 2012.
26. C. Cuevas and N. García, "Improved background modeling for real-time spatio-temporal non-parametric moving object detection strategies," *Image and Vision Computing*, vol. null, June 2013.
27. A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, May 2014.

28. P. Ricaurte, C. Chilán, C. A. Aguilera-Carrasco, B. X. Vintimilla, and A. D. Sappa, "Feature point descriptors: infrared and visible spectra.," *Sensors*, vol. 14, pp. 3690–701, Jan. 2014.
29. M. G. Wing, A. Eklund, and L. D. Kellogg, "Consumer-Grade Global Positioning System (GPS) Accuracy and Reliability," *Journal of Forestry*, vol. 103, no. 4, p. 5, 2005.
30. Z. Zivkovic and F. van der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, 2006.
31. A. Godbehere, A. Matsukawa, and K. Goldberg, "Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation," in *American Control Conference*, pp. 4305–4312, IEEE, 2012.
32. D. Hall, J. Nascimento, P. Ribeiro, E. Andrade, P. Moreno, S. Pesnel, T. List, R. Emonet, R. B. Fisher, J. S. Victor, and J. L. Crowley, "Comparison of target detection algorithms using adaptive background models," in *Proc. Int. W'shop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 113–120, 2005.
33. C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley-Blackwell, 2010. ISBN-13: 978-0470844731.
34. L. Fei-Fei and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories," in *Computer Vision and Pattern Recognition*, vol. 2, pp. 524–531, IEEE, 2005.
35. J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering Object Categories in Image Collections," in *Proceedings of the International Conference on Computer Vision*, 2005.
36. J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object Retrieval with Large Vocabularies and Fast Spatial Matching," in *Int. Conf. Computer Vision and Pattern Recognition*, pp. 1–8, 2007.
37. H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
38. C. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
39. C. Chang and C. Lin, "Training v-support vector regression: theory and algorithms.," *Neural computation*, vol. 14, pp. 1959–77, Aug. 2002.
40. D. Walger, T. Breckon, A. Gaszczak, and T. Popham, "A comparison of features for regression-based driver head pose estimation under varying illumination conditions," in *Proc. International Workshop on Computational Intelligence for Multimedia Understanding*, pp. 1–5, IEEE, November 2014.
41. X. Li and T. Breckon, "Combining motion segmentation and feature based tracking for object classification and anomaly detection," in *Proc. 4th European Conference on Visual Media Production*, pp. I–6, IET, November 2007.
42. W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, pp. 18–32, Jan. 2014.
43. I. Katramados and T. Breckon, "Real-time visual saliency by division of gaussians," in *Proc. International Conference on Image Processing*, pp. 1741–1744, IEEE, September 2011.
44. I. Katramados, S. Crumpler, and T. Breckon, "Real-time traversable surface detection by colour space fusion and temporal analysis," in *Proc. International Conference on Computer Vision Systems*, vol. 5815 of *Lecture Notes in Computer Science*, pp. 265–274, Springer, 2009.
45. M. Breszcz, T. Breckon, and I. Cowling, "Real-time mosaicing from unconstrained video imagery for uav applications," in *Proc. 26th International Conference on Unmanned Air Vehicle Systems*, pp. 32.1–32.8, April 2011.
46. R. Chereau and T. Breckon, "Robust motion filtering as an enabler to video stabilization for a tele-operated mobile robot," in *Proc. SPIE Electro-Optical Remote Sensing, Photonic Technologies, and Applications VII*, vol. 8897, pp. 1–17, SPIE, September 2013.
47. M. Breszcz and T. Breckon, "Real-time construction and visualization of drift-free video mosaics from unconstrained camera motion," *IET J. Engineering*, vol. 2015, pp. 1–12, August 2015.
48. M. L. Eichner and T. Breckon, "Integrated speed limit detection and recognition from real-time video," in *Proc. IEEE Intelligent Vehicles Symposium*, pp. 626–631, IEEE, June 2008.
49. I. Tang and T. Breckon, "Automatic road environment classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 476–484, June 2011.
50. O. Hamilton, T. Breckon, X. Bai, and S. Kamata, "A foreground object based quantitative assessment of dense stereo approaches for use in automotive environments," in *Proc. International Conference on Image Processing*, pp. pp. 418–422, IEEE, September 2013.
51. P. Pinggera, T. Breckon, and H. Bischof, "On cross-spectral stereo matching using dense gradient features," in *Proc. British Machine Vision Conference*, pp. 526.1–526.12, September 2012.
52. M. Magnabosco and T. Breckon, "Cross-spectral visual Simultaneous Localization And Mapping (SLAM) with sensor handover," *Robotics and Autonomous Systems*, vol. 63, pp. 195–208, February 2013.